DOCUMENT RESUME

ED 205 173 IR 009 476

AUTHOR Lofstrom, Mats: And Others
TITLE A Framework for a Future Swedish Policy for Research and Development in Information Science and Technology.
INSTITUTION Royal Inst. of Tech., Stockholm (Sweden). Library.
REPORT NO TRITA-LIB-4069
PUB DATE Mar 81
NOTE 60p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Artificial Intelligence: *Information Dissemination: *Information Science: Information Systems: Library Automation: *Policy Formation: *Program Development: *Research Design
IDENTIFIERS *Scientific and Technical Information

ABSTRACT

Prepared to stimulate discussion on how to design a Swedish policy in information science and technology, this report presents the state-of-the-art of this field as it pertains to the dissemination of scientific information and outlines a program for future research and development. The review portion examines systems for current information retrieval, teleconferencing, office information, and library automation: investigates common problem areas of database design: provides impressions gained from research into artificial intelligence and linguistics: reviews ongoing projects in library automation in the United States: relates current research in Sweden to the discussion of common problem areas: and discusses the need for cooperation among Swedish research organizations. The remainder of the report describes a proposed research and development framework for the Delegation for Scientific and Technical Information (DFI), discusses a suitable research organization by contrasting existing models, summarizes the main proposals for DFI policy, and establishes a time schedule for action. Appendices list research groups working with information processing in Sweden, and contacts used in formulating the report. A 90-item bibliography is included. (RAA)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED205173

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

REPORT TRITA-LIB-4069

MARCH 1981

# A FRAMEWORK FOR A FUTURE SWEDISH POLICY FOR RESEARCH AND DEVELOPMENT IN INFORMATION SCIENCE AND TECHNOLOGY

Mats Löfström, Jan Hultgren and Mats G. Lindquist

Paralog

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stephan Schwarz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR009776

Foreword.

This study reflects our view of the state of the art of information science and technology. In it are also put forth some ideas, based on this view, of how to form a DFI policy for the future.

Our understanding of the field springs partly from the experience gained when developing the 3RIP system, a project sponsored by DFI. Literature studies and visits to some research institutes in the US as part of the present study have added to this knowledge.

The fact that we have tried to cover many subject areas in a relatively short time means that some subjects are treated in less detail than others, partly due to our inital lack of in depth knowledge in certain areas. We apologize to those readers who find the descriptions and arguments too simplified to their tastes.

We hope that in spite of this the report will initiate and provide fuel for a meaningful discussion of how to design a future Swedish policy in information science and technology. To facilitate such a discussion we have tried to be very specific, defining the terms used and including several examples.

The chapters on artificial intelligence techniques and library automation systems have been written by Jan Hultgren and Mats G. Lindquist respectively. They also visited the American organisations listed in appendix B. Mats Löfström has written the rest of the report.

Acknowledgements.

As shown in appendix B many people have been kind enough to devote their time to help us with this report. We would like to thank them all and especially Hans Nilsson, Jacob Palme, and Lars Kahn, who went out of their way to provide us with much valuable material.

Table of contents

1. Introduction 5

2. Descriptions of various system types
   2.1 Information retrieval systems 7
   2.2 Teleconferencing systems 8
   2.3 Office information systems 9
   2.4 Library automation systems 10

3. Common problem areas
   3.1 How to design a data base system and a user interface 12
   3.2 How to integrate different system types 14
   3.3 How to use distributed data base techniques 16
   3.4 How to develop small systems 18
   3.5 How to deal with the problem of size 19

4. Artificial intelligence and linguistics: some impressions from the US 22

5. Ongoing projects in library automation in the US
   5.1 The bibliographic utilities 26
   5.2 Commercially available systems 28
   5.3 Academic systems and research projects 28
   5.4 Summary of current development efforts 29

6. Related projects in Sweden
   6.1 Research work sponsored by STU within the Programme for Research in Information Processing and Computer Science 31
   6.2 Other research and development activities 33
   6.3 Other related projects 34
   6.4 Other organisations of interest 35

7. The reasons for cooperation 36

8. A framework for a programme in information science and technology 37

9. Discussion about suitable organisations
   9.1 How to organise sponsored research work 39
   9.2 How to get useful products from research efforts 41
      9.2.1 The DATALAB example 42
      9.2.2 The 3RIP example 43
      9.2.3 The LIBRIS example 43
   9.3 Concluding remarks on the differences between project types 44

10. Suggestions 46

11. A schedule of time and action 48

App. A: Research groups working with information processing in Sweden 49
App. B: Persons with whom contact has been taken 51
List of references 53

# 1. Introduction

In the present study we discuss a future research and development programme for the Delegation for Scientific and Technical Information (DFI), i.e. what projects DFI should initiate and/or support in those parts of information science and technology (IS&T) that can be said to be within the responsibility of DFI.

In order to make the meaning of the term IS&T clear we need a useful definition and have adopted one used by <Resnikoff 1980>:

> "Information science is concerned with the structure of information and its transfer, i.e. with the ways in which one system is informed by another.
>
> Information technology is concerned with the technical means available for use in effecting such an exchange."

We do not try to break down any possible overall objectives of DFI into their IS&T parts. Instead we merely present a number of system types, that should be considered when a future DFI policy is to be formed, either because their functions could be strongly influenced by actions from DFI or because their evolution will have an impact on the DFI policy decisions. It should also be noted that this study is geared towards the information technology part of IS&T, which means that many of the behavioral aspects of information science are left out together with discussions about the political and economical impact of the techniques and tools presented. In for instance <Riksdataförbundet 1980> we find a survey of ongoing Swedish work in these areas and a study for DFI by <Höglund 1981> discusses some of these problems in the context of the dissemination of scientific information. Finally, we only briefly mention so called "new media" for communication such as viewdata, telefax, electronic publishing, as this subject is meant to be studied separately by DFI.

It should also be noted that, due to time limitations, we have not been looking in any detail into the different IS&T policies adopted by governmental bodies equivalent to DFI in other countries. We believe that much could be learnt from the policies of, for instance, the US, Japan, France, the UK, Germany and Norway and that a survey of different IS&T policies would be worthwhile for DFI. There is in fact a thorough description of the French telematic policy in <Nora and Minc 1980>, which could serve as a starting point for such a study.

The report can be said to be divided into two parts. In the first one consisting of chapters 2-8, we present the arguments for the framework for a programme in IS&T proposed in chapter 8.

In chapter 2 we describe a number of system types. Information retrieval systems, teleconferencing systems, office information systems, and library automation systems are presented, predictions are made about future developments and trends and

the relevance for DFI of each system type is discussed.

Chapter 3 describes problems common to these system types and gives specific examples of problem areas that are interesting from a DFI point of view.

In chapters four and five we report the views we found prevalent among the researchers we met in the US in the fields of artificial intelligence, linguistics, and library automation, while chapter six gives a list of Swedish projects that are in some way related to the system types described.

Why it is important to cooperate within the field of IS&T is discussed in chapter 7 and finally, in chapter 8, all of the foregoing is summarized into an outline of a programme for research and development in IS&T.

The second part, consisting of chapters 9-11, dealing with the problem of how to organise research work, argues that a set of research teams is a suitable way of doing this. The importance of separating research projects from development projects is also discussed. The main suggestions are summarized in chapter 10, while chapter 11 gives a short list of possible actions that DFI could undertake now to accomplish the suggested programme.

## 2. Description of various system types

### 2.1 Information retrieval systems.

By an information retrieval system in this context we mean a computerised system where information is stored in data bases and where relevant facts - be they numeric values, paragraphs of text, or sets of references - about a certain subject matter can be retrieved using a dialog technique, where the user describes a subject matter and refines his description step by step until the system has found and displayed the wanted information. In <Lancaster 1973> a review is given of the basic aspects of information retrieval systems.

The development of bibliographic data bases and services for searching these data bases on a commercial basis have been particularly rapid both in the US and in Europe, as described for instance in <Bourne 1980>, <Regazzi 1980> and <Tomberg 1980>. In the US there is a trend towards the utilization of information retrieval techniques in other areas, particularly in finance and business. In a recent survey <Frost&Sullivan 1980> it is shown that this is also happening in Europe although at a slower rate. This has important implications for DFI as it will strongly affect the basis for all decisions about possible future data base policies and thus also have an impact on the policy for research and development.

Many "new" information retrieval systems will also be developed in the next five years as the market for such systems grows, see <Ashford 1979>, but they will all have the basic characteristics of the present systems. It is not likely, however, that many more systems will be developed on the scale of STAIRS or 3RIP <Paralog 1980> in the future, since the necessary investments are too large to be justfied from a commercial point of view. If large data bases are to be handled by many users simultaneously, a problem which these systems are designed to cope with, the solution will be to try to distribute the data bases among several computer systems or to try to develop systems using new hardware tools. This will lead to new problems, notably dealing with distributed data base techniques, interfacing different retrieval systems, and data base design for special hardware.

As noted in the section about office information systems below, the utilization of these systems by small firms can also bring about a radical change in the demand and supply of information services.

## 2.2 Teleconferencing systems.

A teleconferencing system is used as a means to communicate between people located at different places, as described for instance in <Hiltz and Turoff 1978>. In such a system a user can either send private "letters" to other users of the system or participate in a number of conferences to which he can at any time send any number of "messages". When a user connects to the system he is notified of all unread letters and messages, which he can then scan through. This means that, in contrast to normal meetings, members of a conference seldom participate in the conference at the same time. The fact that you can participate at your own convenience and the savings in travel costs have been cited as major advantages of teleconferencing systems.

The National Swedish Defense Research Institute (FOA) has developed the teleconferencing system COM <Palme December 1980>, which has been in use at the Stockholm Computing Centre (QZ) for several years and is now used by hundreds of people from various organisations. In <Palme November 1980> we find a number of examples of how a system like COM can be used:

> "Exchange of experience between people with similar tasks but placed in different locations (....)
>
> Collection of comments and ideas on suggestions and proposals from a geographically distributed group of people (.....)
>
> Contacts in groups working toward a common goal but spread at different locations, for example, the members of a standards committee.
>
> Getting answers to questions by posing them to a group of people knowlegeable in the subject area of question, for example, how to find a product with particular characteristics."

As can be seen from the last examples a teleconferencing system can in fact be used as an information retrieval system, where questions are put not to a data base but to a number of people. We have successfully used COM in such a way when preparing this report.

If we look at future developments, we believe that there will be a strong demand for integrating different teleconferencing systems with the objective to provide a worldwide system that can be used in various cooperative efforts. Thus the COST-11 project, for instance, decided to develop a machine independent version of COM to be called PORTACOM. Another project within COST-11, called COST-11 bis and dealing with techniques for distributed computer systems is considering to start a project called "GILT - Interconnection between local text systems", where PORTACOM would be used both as a means of communicating within the project and as the basis for research in and development of techniques for transferring messages within a distributed computer system <Palme 80-11-22>.

We also anticipate that teleconferencing systems will play a central role in many integrated systems for communication and that one common request will be to couple teleconferencing systems to other system types. Paralog has already developed an interface between COM and 3RIP in cooperation with FOA <Lindquist M.G. 1980>.

## 2.3 Office information systems.

By an automated office information system we mean a computerised system that attempts to perform the functions of the ordinary office by means of a computer system. Automation, then, would mean to aid an office worker in document preparation, various types of information management such as keeping a diary, and desicion making. Of course the complexity of such a system varies: it can be as simple as a couple of independent word processors or it can comprise complex systems hosted on a series of different computer systems connected through a network. A thorough description of the state of the art of OIS is found in <Ellis and Nutt 1980>.

It is a fact that the market for OIS is growing very rapidly, see for instance <Electronic Business 1980> and <Frost&Sullivan 1981>, the reason for this being that

- office work has not so far been automated to any considerable extent and it is believed that office work could be made more efficient by automation,
- the advent of inexpensive computers and good enough software has made it possible to provide attractive OIS at a reasonable price.

Several Swedish firms are entering the market; LM Ericsson, for example, has opened a new division for information systems and it is believed <Dagens Industri 1981> that a considerable part of LM's turnover will come from the OIS market in 1985.

The fact that most suppliers of computer products, both hardware and software, appreciate that the OIS market is growing fast means that large investments will be made to develop products for this market. The effect of this will be that many of the research problems that we discuss in this study, notably those dealing with artificial intelligence, with the problems of integrating different system types, and with distributed systems, will be looked into from a commercial point of view and that the operational feasability of methods, developed by various research groups, will be thoroughly tested.

Another effect will be that prices for OIS will tend to fall, which means that most small firms will be able to afford to use one. Very many of them will in fact use an OIS, one effect of which will be that the amount of machine readable text will increase drastically and that there will be a large market for small inexpensive information retrieval systems to handle this mostly textual information. Another effect might be that small firms, through getting and getting used to computer equipment and to try to keep a competitive edge, will want to use external

data base services available through networks. If so, this again would mean a large market for microprocessor based systems like those discussed in section 3.4 and, of course, that the market for commercially available data bases will grow as will the demand for special information services that can handle and be a guide to the expected vast supply of external data bases. Such an evolution will also put pressure on the vendors of information retrieval systems as the commercial success of these systems will depend on the relative ease of use for the inexperienced user and on their ability to integrate with other systems in an OIS environment.

To sum up this discussion: we doubt that OIS should be considered to be within the scope of DFI, but the evolution and use of OIS will have a dramatic effect on demand and supply of information services, and thus will have to be followed very carefully by DFI.

## 2.4 Library automation systems.

For the purposes of this study we mean by the term "library automation" computer systems performing internal library functions, e.g. circulation control and acquisition, and systems connecting two or more library units for purposes such as interlibrary lending or the distribution of cataloging records.

Library automation systems first appeared in the early 1970's and their development was dominated by two factors: one was the availability of machine readable catalog records (MARC), the other was the demand for aids for circulation control (being the most tangible of the library functions where automated means could increase efficiency).

The availability of cataloging information in machine readable form represented a large savings potential and many projects were launched to develop systems for cooperative use of the MARC records. The most successful was OCLC whose centralized cataloging support system began operations in 1971. Other large projects were OULCS at the University of Toronto, and BALLOTS at Stanford University. OULCS had a different philosophy than OCLC and instead of having one central common data base the system consisted of a number of "private" data bases, i.e. owned by individual libraries. BALLOTS was an attempt to develop a total library automation system including circulation, cataloging, and acquisition.

The commercial sector concentrated its efforts on developing systems for small library operations, and usually the central function in the systems was circulation control. These systems typically had few or no facilities for telecommunication connections.

After a few years some problems became appearent. In a 1974 study <Lindquist 1974> the following problems were considered

important:

1. There were technical problems relating to character sets and the internal representation of character codes. The requirements from the library community could not be accommodated by the computer industry, hence character representation was made on an ad-hoc basis. For printing purposes the "ALA print train" became close to a standard.

2. The connection to and interconnection of networks were difficult due to the lack of standards for protocols and other procedures.

3. The usefulness of the MARC data bases was limited by the fact that they only contained monograph material and that in some cases the quality of the cataloging information could be questioned (input was sometimes done in a decentralized fashion without any quality control procedures).

4. The increase in efficiency from automation was limited since only one or two of the library functions were automated.

5. User access to the automated catalog was complicated. In particular subject access was not provided.

6. Network development and policy was not coordinated. A considerable duplication of effort took place in the regional network projects.

This list of problems is mainly still relevant.

# 3. Common problem areas

## 3.1 How to design a data base system and a user interface

In this section we will briefly discuss the inherent problems when an information retrieval system suitable for a certain application is to be developed. To begin with we restrict the problem by saying that the system should be run on a single computer and that it should not necessarily be integrated with any other system, i.e. the system should be "centralized and stand-alone", as, indeed, most systems are today.

A system for information retrieval can be said to consist of
- one or more data bases;
- a data base management system (DBMS) by which the data bases can be manipulated in certain ways, e.g. in order to retrieve relevant parts according to certain criteria or to change the contents of the data bases;
- a user process within which the user-system interaction takes place.

To illustrate the magnitude of this problem let us consider what types of data bases might be involved by using a classification scheme developed by Cuadra Associates <Wanger 1980>. They identify the following data base types:
- bibliographic data bases, which contain citations or citations and abstracts from the primary literature;
- referral data bases, which contain descriptions of projects, institutions, activities, the expertise of individuals, and other nonprint/document sources;
- numeric data bases, which contain numeric values from original surveys and/or data that have been summarized or statistically manipulated;
- textual-numeric data bases, which contain records with fields of textual information, numeric information, or both;
- full-text data bases, which contain records of the complete text of an item or some primary source.

Furthermore, the user process will have to be designed so that the requirements of many different user groups are fulfilled. Examples of users might be the general citizen, who is totally unexperienced with computers, and the skilled documentalist or research worker, who needs very little help from the system but wants it to perform many subtle functions.

The problem then is: how to develop an adequate system assuming a given mix of the above types of data bases (which mix may change over time) and expecting it to be used by certain user groups (whose composition may also change over time)? or rather: what are the existing products or tools that can be used in such a system?

We can identify four basically different approaches to the problem:

1. To develop the system from scratch, i.e. to use one of the existing programming languages and code all parts of the DBMS and user process. For a nontrivial system this method is comparatively expensive because of the high cost of programmers and participating users. Another problem might be to maintain the system once it is developed (see the discussion in section 9.2). The advantage is of course that, in the case of a successful development, one gets exactly what one wants.

2. To use a commercially available retrieval system. Being the developers of such a system we would of course like to advocate this approach. However, because of the variety of data base types and user groups mentioned above, it will be very difficult in most cases to find a system that meets the requirements. The advantage of buying a system is of course that this is by far the cheapest way of solving the problem.

3. To build the system using a commercially available DBMS. This approach is a good one for certain applications, e.g. where the data bases are of moderate size and well structured and where searching is performed using simple search criteria. TOTAL and 1022 are good examples of the many available system. Other problems, notably those concerning small data bases and complicated search criteria, may be solved using the relational data model <Codd 1970>. ORACLE, INGRES and the recently announced SQL/DS <Codd 1981> are examples of commercially available systems supporting this model.

   However, many of the above mentioned data bases include textual material, which cannot be structured in the ways required by the above mentioned DBMSs. Other characteristics of textual material are that it often results in fairly large data bases, e.g. containing legal texts, and that the search criteria that one would want to use are not offered by the DBMS mentioned. Commercially available systems such as 1022, ADABAS, and recently MIMER <Altemark 1980> have been tried for applications with text, but no successful such attempt is known to us.

   Finally, we note that the advent of so called data base machines might change this situation radically. This is discussed in section 3.5.

4. To use, if they existed, so called "superroutines". The term superroutine is used for a module (or general purpose program or system) which can serve a class of applications. The idea is that programs that define a specific part of a certain application could be hooked on to the supperroutine. This term was introduced in <Sandewall 1979>, where a disussion of how a superroutine could be realised is found.

   Another similar approach would be to use a programming system, holding a so called "embedded programming language", which could be used to program the application specific parts of the system. CS4 <Janning 1979> is an example of

such a programming system, based om the relational data model. The system is used to develop what are called experimental systems.

The above discussion is of course also applicable to the problem of developing the user process once you have got a DBMS that serves your particular application. In chapter 4 the problems of systems such as LIFER, IIDA and INTELLECT are mentioned. Another example of a generalized tool for dialogue design is IDECS described in <Hägglund 1980>.

To contribute to the best possible solution of the above problem should be one of the overall objectives for DFI, in as far as its involvement in R&D in information science and technology is concerned. It seems to us that this is not so much a question of building or buying systems as one of creating adequate tools for an efficient development of systems. We have not been able to find any commercially available DBMS that is specially designed for the handling of text. Such a system would be most useful, both when developing systems and for research purposes, as such a system could serve as a building block in prototype systems, as for instance in the approach used within the POLYTEXT project (see section 6.2).

We also believe that the above mentioned technique using superroutines or other similar tools could yield a programming environment where it would be possible to create systems that could be tuned to certain applications, e.g. where different combinations of data bases and user groups are involved, in a way that is not possible today. Therefore DFI should sponsor research work in this area.

It should be remembered that we have restricted the discussion to "stand-alone" systems only. The aspect of creating flexible systems is even more important if we require that the developed systems should be able to communicate with other system types. This problem is discussed in the next section.

## 3.2 How to integrate different system types.

Intense discussions, not only in the field of IS&T, is going on about the integration of different system types into one system. By integration in this context we mean "functional integration", i.e. the creation of a system that could be used without the user having to learn the different command languages or other special charcteristics of the subsystems. It should be possible to use the integrated system without knowing that there are, in fact, several subsystems involved. Most people seem to agree that integration is the obvious way to proceed to develop "complete" systems for the handling of information but opions vary on how to achieve functional integration. Furthermore, due to the fact that only a few operational integrated systems exist from which experience may be gained, knowledge about the problems involved is poor.

To illustrate the magnitude of this problem we use a projection by Lancaster <Lancaster 1980>, who foresees that in the year 2001 a research worker will use on-line facilities to:

"(a) maintain notebooks and calendars,
(b) receive and transmit mail,
(c) compose reports,
(d) transmit reports to collegues for informal review,
(e) communicate with editors and referees,
(f) "conference" with collegues working in related research areas,
(g) receive notifications of new electronic publication through on-line services,
(h) search data bases and data banks,
(i) communicate with information specialists,
(k) reguest, receive, and read text and
(l) build and maintain personal information files (the electronic equivalents of laboratory notebooks, diaries, files of reprints, and the like)."

We see that within such a system functions from all of the system types described in chapter 2 have been integrated. In opposition to Lancaster, however, we believe that these kinds of services, except for the functions relying on the existence of electronic publishing, might well be available within the next ten years.

Other examples of functional integration might be:

- a CAD (Computer Aided Design) system to an information retrieval system, where the latter is used to search in a products data base;
- a CAI (Computer Aided Instruction) system to such a library automation system and information retrieval system as is used within the Lister Hill Center <Goldstein 1979> - a report which also illustrates the magnitude of the problems involved in integration;
- the integration of what is commonly called a Decision Support System, where computerised systems for budget simulation and forecasting, for instance, might be integrated with various information retrieval systems used to search in data bases containing economic, business, or legislative information.

The problem of integrating systems can be divided into two parts:

1. Dealing with the problem of integrating existing systems. To solve this problem would mean to define how communication between the systems should be achieved and notably how control should be passed between the systems. One approach would be to define some form of "messages" that could be passed between the systems. As existing systems rarely are designed to interact with other systems, integration must be achieved by re-programming the systems involved.

2. Dealing with the development of systems that are easy to integrate. This problem concerns the definition of a useful interface, i.e. a boundary between two interacting systems

> across which the interaction is fully defined <Davies 1979>, and the design of systems with characteristics that are adapted to this interface. The problem is related to the ones described in section 3.1 where we discussed the development of an information retrieval system.

Much attention is paid to this latter problem in the area of computer science. The advent of a number of operating systems, where parallell processes can be made to cooperate using special facilities for communication, has made the implementation of interacting systems easier. Furthermore, high level languages, such as Ada <Ada 1980>, have been designed with primitives enabling the programming (using what in Ada are called tasks) of systems able to communicate with each other.

This particular research is still in its infancy and it is obviously a very important area that will have to be followed by DFI with much care. We suggest that DFI initiates some prototype versions of functionally integrated systems using existing systems, like the COM-3RIP system, in order to obtain more knowledge about the problems involved.

Note that nothing has been said hitherto about the computer configuration used to run the integrated systems on. Obviously the problems are aggrevated if the systems are to be run in a distributed mode on some network rather than on a stand-alone installation. Something will be said about the particular problems of distributed systems in the next section.

## 3.3 How to use distributed data base techniques.

To achieve "functional" integration in the sense defined in section 3.2, when the systems involved and their associated data bases are spread among different computer systems, one has to use what is called distributed data base (DDB) techniques.

We define DDB using <PACTEL 1979>:

> "A distributed data base exists where logically related data is physically distributed between a number of separate information processors linked by telecommunications.
>
> Information processors are data processing computers on which application programs, or parts of application programs, can be loaded and run. (.....) There must exist some form of logical connection between the data held at separate information processors, established and maintained by means of some system or application process."

To try to describe the many problems involved in DDB and distributed processing would take us too far, but we refer the reader to <PACTEL 1979> for an introductory description of the problem area and for discussions about future trends. An extensive list of references is also included there.

We will just note here that the incentives for using DDB

techniques in future systems for information handling are many:

1. We have already noted the direction towards functional integration of systems. When the systems are spread among various computers in a network, the lack of appropriate DDB techniques may inhibit the development of an otherwise desirable integrated system.

2. Certain applications are best suited for a decentralized solution, where for instance certain data bases are controlled locally. Such an approach can be realized using DDB.

3. Under certain conditions the reliability of a distributed system is much better than that of a centralized approach.

4. A DDB approach might offer a good solution to the problem of how to handle a growing population of terminal users of, for instance, a large information retrieval system.

There are of course many other motives for trying to develop DDB techniques but the above list will suffice to point out that this area is and will be exceedingly important for DFI to follow and that research efforts in this direction should be supported.

It is also important to point out that the success of a DDB approach to a certain problem relies not only on the solution to the technical problems but on other factors such as:

- the development of standards, not only for communication protocols such as the six levels of the ANSI/X3/SPARC reference model, but also, for example, for the data bases themselves. There are no recognized data base standards and until such standards emerge, if they ever do, translation will be necessary for all data base commands and for all control commands of the various computer systems involved, making the work of developing distributed systems much harder;
- whether the organisations that control the various data bases and systems can cooperate successfully. It will be necessary to create an adequate forum for all parties interested in the development of a system using DDB techniques.

Finally, we note that using DDB techniques means, possibly more so than in any other area of IS&T, to approach intricate technical problems. It is therefore mandatory for an organisation that perceives a demand for DDB techniques in the future to start now to build up a competence in the field and try to see to it that this competence is kept within the organisation.

## 3.4 How to develop small systems

With the advent of small, inexpensive and yet powerful microprocessor based computer systems, there is a great interest in how to use these tools in various fields: as a stand-alone system for a certain application, for improving the performance of an existing system, or for customizing it to the needs of a certain user or user group. In this section we will use the term "small system" for all systems that are based on these inexpensive computers and that are designed to be used by one user at a time, i.e. the systems should not be extended into multi-user systems.

To illustrate how a small system could be used in the field of information retrieval, we use an example from <Williams 1980>, where the problems of developing a small system to be used as an aid for searching on an external information retrieval system, e.g. Lockheed or BRS, are described. The system is used both to avoid high telecommunications cost and for customization of the user-system interaction to a certain user group. Williams recognizes the following improvements that a user might want:

> "(i) automatic log-on;
> (ii) storage and editing of profiles before transmission;
> (iii) storage of profiles and frequently used command sequences;
> (iv) re-formatting of output to local requirements;
> (v) selecting from output and annotating with personal comments;
> (vi) automatic collection of statistics on systems used and charges incurred;
> (vii) provision of alternative dialogues for new and skilled users;
> (viii) ability to switch freely between different systems;
> (ix) conversational query systems for unskilled users; and
> (x) systems using users comments to improve the search profile."

Assuming that it is possible adequately to specify what a small system should do the problems of developing such a system are:

1. to find a suitable small computer among the many existing ones and

2. to program the system in such a way that one gets as many facilities as possible into the system. Today the programming is performed in one of the high level languages (usually BASIC or Pascal) that is available on the particular computer.

We would say that of these the first one constitutes the greater problem, at least for the inexperienced user, as the market for both hardware and software for small systems is somewhat overheated, resulting in dubious products sold at a very low price. One could well get a product that is unreliable, full of so called "bugs", seriously hampering the system development.

To solve the second problem is usually a straightforward

programming task given the limitations of the computer system and the programming language chosen, always supposing that one does not try to develop a so called portable system, i.e. a system that could be moved between various hardware systems without extensive reprogramming.

Thus we find that there are good reasons to develop various small systems to facilitate the utilization of information handling systems. We may also note that it will be possible to put many more facilities into a future small system since the price/performance ratio of hardware is falling and because of the fact that it will be both technically and economically feasible to connect secondary memory devices with increased capacity to small systems.

The problem for DFI is rather to decide whether such development efforts should be supported or not. Good reasons for support would be the decreasing costs and added facilities resulting from the development. The obvious reason against support is that a developed system is not likely to be used outside a small group since it is by nature "personal". There are of course examples of small systems in the sense defined above, such as SOF <SOF 1981>, that is claimed to be portable and thus useful for a large group of users but we do not find many successful such attempts.

When generalized and reliable tools for systems development, such as the programming system Rosetta Smalltalk <Warren 1979>, are available on small systems the situation might change so that it will in fact be possible to develop portable systems.

For the present we would suggest that the most important thing for DFI is to see to it that competent groups exist that could guide potential users or developers of small systems to the vast supply of available hardware and software products. For this reason it is important to support some development efforts from which experience could be gained of how to use these products in practice.

### 3.5 How to deal with the problem of size

Many applications within the field of IS&T require that a system should have a high capacity, i.e. that many users should be able to interact with the system simultaneously and still get acceptable response times and also that the system should be able to handle very large data bases. The latter requirement will become more accentuated in the future as for example various large archives are being computerised. We have also noted that the amount of machine readable information will increase drastically in the future, because of the production of textual material from, for instance, office information systems and also because of various OCR devices like the KDEM. There will probably also be a demand for systems that can handle both textual and pictorial information (see for instance the discussion about a future Swedish national encyclopedia in <SOU

1980:26>), requiring that very large data bases can be handled efficiently.

We have already mentioned that DDB techniques might be used in certain cases to solve this problem of size. In this section we shall introduce two promising tools that might solve some of the problems. They are the videodisc and the data base machine.

When talking about videodiscs one should differentiate between two types <Goldstein 1979>:

1. the "optical videodisc", on which a normal videosignal is recorded; this type of disc can be replicated at a very low cost by "stamping" it from a master disc;

2. the "optical digital disc" <Bartolini 1978>, which is used to store and retrieve digital and textual information, which is recorded at the computer site, i.e. there is no mastering or replication process in this case; this type of disc is used in various systems for mass storage <Ammon 1979>.

Both types of discs allow for relatively fast random access to large amounts of information: up to 10 thousand million characters have been recorded on one side of an optical digital disc. Furthermore, the lifetime of a disc is almost unlimited so that a videodisc is well suited for archival purposes.

The possible applications of videodisc technology are many. They could be used:
(a) in systems where both textual and pictorial information is to be stored and retrieved;
(b) in various hardware configurations as a device with access times between those of magnetic tape and magnetic discs and with the possibility of archival storage;
(c) to replicate masses of information at low cost; for instance, large data bases could be replicated and then distributed to a number of computer sites where they could be handled locally.

Before the videodisc could be used in an operational environment a device that controls the disc and with which a computer system could interact is needed. Moreover, all of the above mentioned applications presuppose the solution to various intricate technical problems which are being approached in research laboratories. Thus:
- application (c) requires that digital information can be encoded onto the videosignal and that techniques exist that considerably reduce the random errors introduced during the present mastering and replication process;
- (b) means that new trade-offs are introduced in the design of systems for secondary storage and that new data base techniques might be needed, see <Copeland 1980>;
- (a) requires the design of new software systems suitable for the storage and retrieval of textual and pictorial information mixed together on one storage medium.

The objective of the data base machine (DBM) is to implement many of the traditional data base management functions directly in the hardware to increase performance and capability. Two basic reasons lie behind this evolution:

1. As hardware costs are decreasing rapidly while the costs of software development and personel are increasing one wants to find solutions to data base management problems that take maximum advantage of hardware architecture and technology.

2. It has been shown that certain functions of a specialized designed hardware configuration, e.g the possibility to perform several functions in parallel and to use content-addressing <Foster 1976>, are very useful for the processing of non-numeric information. (By content-addressing we mean that a cell is adressed on the basis of its contents rather than on the basis of where it is located in memory.)

An overview of the subject of DBMs together with a list of references is found in <Berra 1977>.

There are a number of possible applications of DBMs within the field of IS&T, the problem again being that there are only a few DBMs commercially available today which would improve the performance of a system to any considerable extent. There are, for instance, machines dedicated to information storage and retrieval, chiefly text processors, which one would imagine would solve the text data base problem discussed in paragraph 3.1, e.g. the high speed text search system <HSTS 1980>. This system performs in hardware most search functions of an advanced information retrieval system such as string searching and boolean logic. The response time from the system is, however, seldom lower than 10 seconds even when searching a data base of moderate size, which makes the system practically useless in any interactive application.

Berra, however, believes (see <Berra 1980>) that in 1985 DBMs will be available that have incorporated new technological breakthroughs so that their performance will be acceptable. If this comes true it will of course have a profound effect on the development of all the system types discussed in this report.

The same could also be said about the evolution of the videodisc technology. To sum up, we believe that the development of data base machines and devices based on videodisc technology must be followed carefully by DFI, for the advent of a certain machine or device might mean that an inexpensive solution to a certain problem is available or the effect might be that much previous work on software solutions might be wasted.

## 4. Artificial Intelligence and Linguistics: some impressions from the US

Various desired characteristics of IR systems, office information systems, etc. fall within the scope of artificial intelligence and linguistics, which disciplines have given some considerable attention to these problems. A few of the subjects that come to mind immediately are:

1. natural command languages in which to adress the machine;
2. facilities for speech input and output;
3. facilities to detect and automatically compensate for trivial mistakes in the input;
4. facilities to accept and resolve ambiguities and fuzziness in the input, if necessary via a reasonably intelligent and helpful dialogue;
5. facilities in the machine for knowledge acquisition and inference drawing;
6. question answering knowledge based systems, requiring new structures for the representation of knowledge, preferably language independent;
7. facilities to alleviate or automate the generation of new vehicles for communicating with the machine, for the teaching of it, and for the planning of its activities.

The efforts put into the research and development in this area, taking these words in the senses used elsewhere in this report, have already produced a number of projects and implementations, wellknown in the literature. The following list is very far from being complete, but the bibliographies in the references cover the field fairly extensively:

1. pattern recognition systems like the speech recognition projects HARPY and HEARSAY II <Lea W.A., Ed. 1980> and MIKE <Centigram>, and the OCR system KDEM <Kurzweil> (MIKE and KDEM also have synthesized speech output facilities);
2. natural language processors like PHRAN <Wilensky R. and Yigal A. 1980>, Linguistic String Project <Sager N. 1978>, LIFER <Hendrix G.G. 1977>, S.P.I.R.I.T. <SYSTEX>, and INTELLECT <AIC>;
3. knowledge acquisitional and representational systems like NANOKLAUS <Hendrix G.G. and Haas N. 1980>, DENDRAL <Lindsay R.K. et al 1980>, MYCIN <Shortliffe E.H. 1976>, and CONDOR <Siemens 1978>;

4. pre- and post-processors to traditional IR systems like CITE <Doszkocs T.E. and Rapp B.A. 1979> and IIDA <Meadow C.T. 1979>;

5. planning tools like META-PLANNING <Wilensky R. 1980>, KBMS <Wiederhold G., Kaplan J.S., and Sagalowicz D. 1980>, and ROSIE (personal communication with Hayes-Rooth F., Rand Corp. Santa Monica, CA. 1980-02-02).

It will be appreciated that the purpose of the rough and ready categorization above is to give a quick overview, but that it oversimplifies matters and that most of the projects/systems do not readily lend themselves to it. The whole point about PHRAN, for example, is that it utilizes knowledge base techniques in its approach to natural language and certainly there is some inherent aspect of natural language analysis in most of the systems mentioned.

One thing that can safely be said about those systems is that they answer the problems enumerated above to a limited extent only, at least if we are asking for developed tools. In order to ascertain what the limitations may be and what period of time it may take before they have been lifted and the research systems have been promoted into full-fledged production systems, we have interviewed a number of people responsible for some of the given projects and also some people in industry (see the list in appendix B).

The following paragraphs summarize the insights gained and the prejudices confirmed. We do not always give credit where credit may be due; conversely it should be understood that not all our informants may wish to agree with everything said in this context. Generally speaking industry is, predictably, more optimistic than research.

MIKE, INTELLECT, S.P.I.R.I.T., etc. exemplify that commercially viable speech systems and query-answering systems exist, but only for small, well defined universes using a restricted vocabulary and a restricted syntax.

The DARPA speech-understanding project <Medress 1978> had as some of its performance goals that the systems created under it should:

- accept connected speech,
- from many speakers,
- permitting a slightly selected vocabulary of 1000 words,
- with a highly artificial syntax and
- a highly constrained task.

At the present time no system does more. The commercial systems do considerably less: connected speech, for example, is not accepted (pauses must be given between the spoken words).

The various new techniques for knowledge acquisition and representation have led to the creation of extremely useful systems for the solving of well defined and constrained problems, but the attempts to cover wider information retrieval tasks are still very much at the laboratory stage. Siemens'

ambitious CONDOR project, with planned facilities for automatic indexation and knowledge acquisition for instance, is a good example of such an attempt, which certainly is meant to result in a competitive commercial product.

As one of the reasons for pessimism is that the linguistic theories do not yet allow finely enough tuned models for the pruning of the tree of possible paths in the unrestricted applications, it follows that one has little enough hope for the more far-reaching projects of natural language analysis to come up with complete answers in the near future.

It will most likely be 20 years or more before we can use connected speech for multi-purpose input or a fact retrieval system with anything approaching encyclopedic knowledge of a broad range of subjects. Professor Reddy takes a more optimistic view - mainly because he expects a greatly increased flow of money and effort in this field. In this context we would like to suggest that the interesting fact about MIKE, KDEM etc. is not how little (or much) they do but rather that they are commercially available.

It must be pointed out here, as it was universally pointed out to us, that the very rapid development in hardware may change the predictions dramatically. However, such changes will only materialize if there is a readiness to match the new hardware with a corresponding development of software techniques. It would be foolish to instigate a major software project without taking a long and hard look at the particular aspects of hardware mentioned elsewhere in this report.

The relative pessimism when it comes to the general, unrestricted problems must not be allowed to hide the usefulness of the great steps forward which have been taken during the 70's. It is felt that in many applications the user looking for information is really only interested in a restricted universe, whose boundaries he has valid preconceptions of. In such cases the existing systems promise to solve most of the problems.

However, it is still the fact that these systems require a considerable amount of software engineering:

- better command languages to enable the end user to create and update such restricted universes;
- better heuristic control to cut down on the resource requirements;
- better planning instruments to administrate and integrate such "island" universes.

The existing techniques are sufficiently stable to allow much of this work to be done as development projects in our sense. One important point, stressed by several people, is that all available sources of knowledge must be utilized; the lack of universally accepted models in many of the relevant areas makes it unlikely that a stubborn pursuit of only a few lines of approach will lead to success.

As we cannot expect radically different systems to appear in the near future the general opinion has it that various schemes to

further develop and improve existing IR systems should be kept going in parallel with the more ambitious research outlined above. Two aspects in particular have been stressed by our informants: how important the problems of administration and organisation are for an information service; how much can be done with existing AI and linguistic techniques to create useful pre- and post-processors to the current IR systems.

## 5. Ongoing projects in library automation in the US

### 5.1 The bibliographic utilities

For on-line access to cataloging information four organisations have emerged as established "bibliographic utilities:" OCLC, UTLAS, RLIN, and WLN. Of these OCLC is the largest having more than 3000 computer terminals in a network of some 2200 libraries. The central MARC data base contains over 6 million bibliographic records. Organisationally a structure where libraries are connected to the OCLC network via "service centers" have evolved (Robinson 1980):

> "OCLC took a laissez-faire approach to the marketing of its services and left it up to libraries to determine their pattern of affiliation. By May 1980, there were 21 OCLC service centers which handled 90% of the terminals linked to OCLC.
>
> Many of the service centers were already well-established state or regional (multistate) networks. A few (...) were formed specifically to broke OCLC's services, but most OCLC service centers also provide a variety of other non-OCLC services to libraries in their network."

The other three networks will be described briefly by quotations from the same JASIS article (Robinson 1980).

UTLAS is the University of Toronto Library Automation System and is the dominating network in Canada:

> "Second in size to OCLC, UTLAS believes in the 'maxi/mini' approach, i.e., that there is a role for centralized computing as well as locally operated minisystems. It offers its users data-processing services and will run a user's file against four data bases within its own files: the MARC tapes, the data base of the National Library of Canada, Bibliotheque National de Quebec, and the National Library of Medicine. A major strength is its software which provides users much flexibility in producing COM catalogs, union lists, cards and tapes. Unlike OCLC, UTLAS is not maintaining a master record data base, although it encourages customers to share one another's files."

RLIN - the Research Libraries Information Network - is run by the Research Libraries Group, Inc. (RLG):

> "RLG is responding to demand for better quality control by abandoning OCLC's notion of a 'melting pot' data base and broad customer base. In contrast to OCLC, RLG has been drawing from a selected group of research libraries across the United States. (...)
>
> Rather than develop its own software, RLG adopted Stanford University's BALLOTS system (...) for cataloging, acquisitions, fund accounting, and serial control. RLG views RLIN as the vehicle for carrying out its programs in the areas of preservation, shared collection development and

management, and shared resources among the research libraries."

The Washington Library Newtork (WLN) still has a regional orientation and has at present no ambitions for developing a nationwide data base. WLN is, however, an alternative for libraries in northwestern USA:

"In addition to online cataloging with subject access, WLN also offers an acquisitions subsystem with fund accounting. It has placed emphasis on quality by offering authority control online as well as manual editing of shared cataloging records by WLN staff before entry into the data base. This kind of control, coupled with subject access, has made WLN attractive."

As can be inferred from these descriptions some of the problems of the early 1970's have been addressed and more or less solved:

- network connection is technically easy, although special terminals are sometimes required, and the networks can handle a large number of terminals;
- quality control of the bibliographic information is emphasized and procedures for the control have been worked out;
- the usefulness of the data bases have increased since serial records are now included;
- authority files online have made it somewhat easier to access and search the data bases.

At present the further development of the large networks follows three avenues: the first is toward increased scope of the service in terms of the number of library functions supported. OCLC's development schedule is representative:

1. cataloging (since 1971);
2. serials control (since 1975);
3. interlibrary loans (since 1979);
4. acquisitions (expected 1981);
5. circulation control and remote catalog access (planned);
6. information retrieval and subject access (planned).

A second line of development is toward greater ease of use of the catalog. WLN's online authority file is an example, and OCLC's plans for information retrieval indicates a higher ambition. UTLAS has an online PRECIS system and is also experimenting with improved subject searching functions for "patron access." In the section on commercial systems below a further example is given.

The third major line of development is toward the interconnection of local systems and the major networks. This aspect is expressed most clearly by UTLAS. Regarding the interconnection of the networks some studies have been made.

The costs and benefits of linking the bibliographic utilities was recently studied under a grant from the Council on Library Resources, Inc. (CLR) <Smalley et al. 1980>. WLN and RLIN are considering some kind of link between the two networks.

## 5.2 Commercially available systems

The commercially available systems are still typically made for one local library even though it is usually possible to connect several installations of a system to each other. The basic library function is still circulation control.

The trend is toward the inclusion of more library functions in the systems (Geac, CLSI, Dataphase, CLASSIC). In particular increased subject access is being emphasized. CLSI considers "public access" to be "the next phase of library automation" and markets a browsing terminal with a touch-sensitive display screen for direct interaction with the users. The terminal enables the system to respond to the patron's touching of the screen to indicate the portion of the data base which she or he wishes to examine in greater detail (Cherry 1981). Geac has announced a "Public query" module to their library information system which allows searching by title, author, call number, or subject.

The acquisition function seems to be the next major function to be supported by the commercial systems.

Another direction of development is interconnections with the major bibliographic utilities. CLSI, for example, offers a module that makes it possible to access the OCLC data base and transfer bibliographic records to the local system.

The commercial systems typically have their own internal formats which sometimes look like LC's MARC format but that sometimes are very different. The trend today is toward complete MARC support (Dataphase, CLASSIC; Geac plans MARC support), i.e. to make the systems work with complete MARC records internally.

## 5.3 Academic systems and research projects

Probably the most complete library automation system today is the NOTIS system at Northwestern University. The system has modules for circulation, cataloging support, acquisitions, and serials check-in, and it supports the MARC format. NOTIS runs on a dedicated IBM computer under the CICS monitor. At Northwestern the symbolically significant step of closing the card catalog was taken in early 1981.

At the Lister Hill Center for Biomedical Communication of the National Library of Medicin an integrated library system (ILS) has been developed which will run on PDP/11 computers, the IBM series 1, and the Data General Eclipse series computers.

The design goals for ILS are:

1. integration of library functions;

2. minicomputer capability;

3. maximum transportability;

4. sizing dependent on hardware only;

5. system network access;

6. multilevel user interface;

7. MEDLARS III compatibility.

ILS Version 1.0 has just been released, and is available on magnetic tape from NTIS for a licensing fee of $2,000. The status of the Lister Hill Center as a government body precludes them from marketing the system and from making maintenance commitments. Based on the ILS, however, a commercial version has been developed by and is available from Online Computer Systems, Inc. The commercial ILS has a minimum price of $50,000 excluding monthly maintenance and updates.

The basic capabilities available in the ILS Version 1.0 are: Master Bibliographic File, Circulation, and Check-in. Present development efforts are directed toward implementing authority files and online catalog access for the reference librarian and patron. Different designs are being tried for the search facility, such as natural language input. Other planned versions of release 1 include reserve book function, multi-department (institution) function, and generalized patron search facility. Plans for release 2 include the acquisitions and cataloging subsystems.

For network interconnection a "black box" is being developed in cooperation with OCLC which will enable linking between networks and local systems. The basic concept is to provide an inexpensive interface that would permit the user to access multiple systems from a single terminal. The purpose of the multiport "black box" is to emulate the data communications protocols unique to the system being accessed rather than relying on a specific terminal to perform that function. The black box is to be programmable so that new protocols can be accomodated.

## 5.4 Summary of current development efforts

Developments in the design of computer terminals and photocomposition equipment have improved the possibilities to produce printed output, such as catalog products, with large character sets. The equipment needed is still too expensive to allow high quality printing at small local sites. Thus, in spite of efforts to decentralize printing, reliance is still on

the major bibliographic utilities.

A major source of information imcompatibility is format differences. The MARC format as used by the Library of Congress is the accepted standard but some systems use subsets of the format or have specialized internal formats. This is especially true about the small commercial systems. There is, however, a trend toward support of the full MARC format.

The major bibliographic utilities as well as the commercial systems vendors continue the development of their systems to include and integrate more of the basic library functions.

The problems relating to network access and interconnection are being studied from both a technical and an economic viewpoint. The most general technical projcet is the "system network access project" at the Lister Hill Center. The economics of linking bibliographic utilities has been studied with funds from CLR. This study can serve as an illustration of the increased attention given to questions of "information economics." The National Science Foundation (NSF) has, for example, given priority to this field under the lable "Infometrics." The NSF program element, however, has a wider scope and "is concerned with the development of models, analogous to econometric models, which will make it possible to predict as well as assess the impact of technological, regulatory, and economic change on the information component of society" <NSF 1979>.

Large efforts are being devoted to make access to the data bases easy for the inexperienced user. In all three sectors surveyed (the major networks, the commercial systems vendors, and the research sector) different approaches are being tried out for subject searching facilities. Perhaps the most ambitious project is the one at the Lister Hill Center where methods for natural language analysis will be used.

# 6. Related projects in Sweden

In this chapter we give a brief overview over some projects that are related to the problem areas discussed. We do not claim that the list is complete but it can still be useful as a guide to ongoing research in Sweden and also serve as a starting point for discussion about possible joint research efforts where DFI might be involved.

## 6.1 Research work sponsored by STU within the Programme for Research in Information Processing and Computer Science.

Most university research projects in information processing and computer science are sponsored by the National Swedish Board for Technical Development (STU) and form part of STU's overall research programme in information processing and computer science, that began in July 1980 and is supposed to last for five years with a total cost of some 80 million SEK <STU 1980>. According to an investigation made for the Office of Swedish Universities and Colleges of Higher Education (UHÄ), cited in <STU 1980>, there are some thirteen research groups dealing with computer science existing in Sweden. These groups are listed in appendix A.

Of the ongoing and planned research projects listed in <STU 1980> we have selected what we believe to be the most interesting ones from a DFI point of view. They are:

1. "A computer network for the research environment."

   The ambition of this project, led by Hans Nilsson, is to "connect the seven large university areas (i.e. Luleå, Umeå, Upsala, Stockholm, Linköping, Göteborg and Lund) into one network such that files can be exchanged. A later level of ambition is access to a mail and conferencing system. In the long run it is estimated that computer independent process-to-process communication can be established.

   The proposed technical solution is a package switching network according to the SWENET solution which is supported by the Swedish telecommunication authority. The exchange protocol will be the standardized X.25-interface."

   If DFI considers to sponsor research and development in distributed data base technology then this research group constitutes a major source of knowledge about and competence in the problem areas involved. Furthermore the planned network provides the first levels of a layered computer network architecture on top of which protocols for the cooperation between, for instance, different information retrieval systems can be built. The network could also be

used as a testbench for different approaches to the problem of connecting personal microcomputer based system to various network services.

2. "A software systems research center"

This research center in Linköping, headed by Erik Sandewall, has a laboratory dealing with

> "application software systems, oriented towards techniques which make computers easier to use for general citizen, particularly techniques for dialogue between the user and the computer system (man-machine dialogues) and for modeling (=description) of applications in the computer."

To test tools and techniques the laboratory is working with reference installations, notably office information systems and systems for hospital and university education. The laboratory addresses many of the problems mentioned in chapter 3, using artificial intelligence techniques, and much could be gained if a DFI sponsored research group could cooperate with this laboratory.

3. "A laboratory for research in programming methodology and artificial intelligence."

This laboratory is a joint effort of the Institution for Information processing at Upsala, the Computer Science Laboratory, and Upsala Computer Center (UDAC). Among the various ongoing projects could be mentioned:

- one project, led by Risch, that attempts to use a meta-data-system for advanced communication between humans and a data base system,
- a group dealing with natural language text comprehension and
- the work going on using the relational data base system MIMER <Altemark S. 1980> in a distributed data base environment and the testing of the applicability of the information retrieval system MIMIS <MIMIS 1980> to library automation work.

4. "A center for research in systems development and software engineering (SYSLAB-S)."

This laboratory is established at the Department of Information Processing and Computer Science at the University of Stockholm and the Royal Institute of Technology and is headed by Janis Bubenko. Within this center one project (MAD) has as one of its objectives to investigate the possibility of creating interfaces between large scale general data base management systems and micro computer applications and is therefore interesting in this context.

5. "A center for industrial computer systems."

This center is a joint effort between the departments of

Automatic Control and Computer Engineering both at the Institute of Technology in Lund. The center hosts a hardware design laboratory that has the ability to develop some of the special hardware components discussed in section 3.5.

6. "A laboratory for research in distributed information processing systems."

   This laboratory is established at the Chalmers University of Technology with Sven Tafvelin as its scientific leader, and is of course interesting as a major source of knowledge about distributed data base systems.

## 6.2 Other research and development activities

Here follows an unstructured list of some interesting projects:

1. DATALAB (see section 9.2.1)

2. The development of the PORTACOM system by ENEA DATA on a FOA contract and the international activities within the COST-11 bis project (see section 2.2)

3. The activities of FOA 4 in Umeå, where techniques to develop expert systems are explored.

4. The work by Paralog to further improve the 3RIP system notably the introduction of the Common Command Language (CCL) <Larsson R. 1980> and the interface to the teleconferencing system COM developed in cooperation with FOA <Lindquist M.G. 1980>. 3RIP is also used in various linguistic applications.

5. The research and development work by KVAL with the POLYTEXT system <Karlgren H. and Walker D. 1980>, <Löf S. 1980>, where artificial intelligence techniques are used in a new approach to represent legal text and to search the material using natural English statements. Furthermore the work will include the interface of an INTERLISP <Sandewall E. 1978> programming system to a data base system (MIMIS and 3RIP have been mentioned) on the one hand and with a system to create user interfaces (also programmed in INTERLISP) on the other. The idea is to build a programming environment where it can be tested whether artificial intelligence techniques can be successfully applied to certain classes of retrieval problems.

6. Activities within Office-85 <Thorngren B. and Eckerström A. 1979>, which is a joint venture between Televerket and major Swedish companies being users of office services rather than manufacturers of equipment. Field tests are carried out using existing word processors and communication equipment to test new facilities in a real office environment.

7. The LIBRIS activities (see section 9.2.3)

8. Kathleen T. Bivins' work on very small IR systems at UCLA and Linköping University.

9. Much of the research at the department of computational linguistics at Gothenburg University (Språkdata) has a bearing on the areas in question.

As noted in section 2.3, vast research investments are made by firms like LM Ericsson, Datema, and Industrimatematik.

## 6.3 Other related projects.

Various research groups and projects are dealing with models of how to evaluate information services. Here we can mention the INFORSK group at Umeå University and an ongoing project led by Anders Sandberg dealing with office information systems <Sandberg A. 1980>.

The evaluation of how information services could be used by the administration, as carried out by the Swedish Agency for Administrative Developments (SAFAD), is of course important for DFI, especially in the context of a possible future large Swedish computerised information service. The plans of SAFAD are presented in <SAFAD 1980:13> and <SAFAD 1980>.
The same argument also holds true for the National Central Bureau of Statistics (SCB).

We should also mention the ISOK-project <ISOK 1980>, which is a project to be carried out by Kommunförbundet, sponsored jointly by STU, Byggforkningsrådet and Delegationen för Transportforskning. The project is an attempt to integrate a number of information systems, including one for the representation of maps, into a complex system for decision making. Much can certainly be learnt from this attempt to implement an integrated system.

Finally, it can be noted, that we have been asked by several people during the work with this report whether we think that the group dealing with the subject of Informatics at the Department of Information Processing and Computer Science at the University of Stockholm is of interest to DFI. Our answer is that from a research and development point of view this group has little to offer since it deals mostly with general systems theory <Samuelson K. 1980> and handles few problems associated with information processing and computer science.

6.4 Other organisations of interest.

When the possibility of getting joint sponsorship for projects is discussed the following points may be of interest:

1. that STU, besides sponsoring the above described progamme for information processing and computer science, also supports other projects within the area of computer science;

2. that the Swedish Council for Planning and Coordination of Research (FRN) finances certain expensive equipment used within research projects sponsored by STU or by Swedish research councils;

3. that the Swedish Organisation for Computer Users (RDF) <RDF 1979> exists. RDF's overall objective is to take care of the interests of users of computerised tools. The organisation also tries to facilitate cooperative efforts when computer products are being developed; thus STU cooperates with RDF to get various forms of industrial support for the R&D programme.

## 7. The reasons for cooperation

As noted in chapter 6 there are many research groups and organisations working with problems related to those described in chapter 3.

We believe that there are mutual advantages for DFI and some of these groups to cooperate, in order to achieve the best possible effect of the money spent on R&D in information science and technology:

1. As several of the groups have been in operation for many years a competence has already been built up that can serve as a basis for DFI sponsored R&D efforts.

2. The field of IS&T is so broad that only a small part of it can be handled within DFI's limited budget. In order to be able to work with larger problems cooperation with other groups is necessary.

3. For many of the existing research groups presently not directly involved with the field of IS&T an extension in this direction should be of interest because
   - the field provides many possible reference applications, e.g. viewdata or archival systems, with user groups willing to participate in the evaluation of the techniques used in for instance a prototype system;
   - it addresses many practical problems that are interesting from a research point of view, e.g. how to use videodiscs and data base machines in information retrieval systems.

4. Competent research groups specializing in the field of IS&T should have much to offer other research groups in terms of solving problems associated with, for instance, the integration of different system types and the handling of textual information.

Thus the reasons for cooperation are strong; the problem is how to achieve it. (We restrict the discussion to cooperation at the national level. There are examples of successful international cooperation, e.g the POLYTEXT project mentioned above, a joint effort of KVAL and Stanford Research Institute.).

There are basically two ways of trying to get existing research groups to work with IS&T. The first one is to market DFI to the university community as an organisation that sponsors research in IS&T, thus trying to make groups apply for research grants from DFI. (It has been noticeable during the work with this report that the existence of DFI - and its predecessor SINFDOK - is practically unknown to the academic community.) The second approach is to organise DFI sponsored research groups in ways which make cooperation possible. This latter approach, which is to be preferred as it has a number of other advantages, is discussed in detail in chapter 9.

## 8. A framework for a programme in information science and technology.

We have so far identified relevant areas for research and devolopment in information science and technology and discussed tools and techniques to solve specific problems. In this section we sum up this discussion into a framework for a DFI Programme for Research and Development in Information Science and Technology. By the terms long, middle and short range, we mean that a proposed activity is expected to last ten, five, and two years, respectively, before it yields any results.

We propose that DFI should support research in information science and technology that adds to the knowledge of:

1. how different system types can be functionally integrated;

2. how distributed data base techniques could be used to achieve functional integration;

3. how various existing and forthcoming programming techniques could be used to develop systems that could be functionally integrated;

4. how artificial intelligence techniques and linguistics could be used to improve the performance and ease of use of systems;

5. how a data base model for textual information should be specified.

1 and 2 are long range activities while 3, 4, and 5 are middle range.

DFI should support development of:

1. certain small systems for information handling; this is a short-range activity,

2. a data base system for textual information; this is a middle range activity and it assumes that activity 5 above provides useful theoretical results.

Furthermore DFI should closely follow the development of:

1. new hardware devices such as data base machines and devices based on videodisc technology;

2. commercially available tools for distributed data base systems;

3. commercially available systems that use artificial intelligence techniques and linguistics to facilitate the interaction between user and system.

We have also been asked to provide estimates on what resources, in terms of manpower, money, and machines, are needed for such a program. We note that, for instance, the SIRIUS project <Bihan 1980>, working with many of the above proposed activities, has an annual budget of more than 70 million francs and occupies about 100 people and that the CONDOR project has occupied more than 50 people within various departments of the Siemens company since 1975 and is expected to go on doing so for the next five years. The objective of CONDOR is to develop a commercially available information retrieval system, where techniques from artificial intelligence, linguistics, and statistics are incorporated into the system.

Obviously these activities require a continuous flow of many and great resources. It is to be hoped that DFI's budget for research and development will be increased in the future. Furthermore, we argue that in order to provide useful results from DFI activities it is necessary that cooperation takes place between DFI sponsored research and various of the research groups mentioned in chapter 6. How this is to be achieved is discussed in the following chapter, where we also argue that the detailed planning of resources should be made in cooperation between DFI and the research teams that are to work with one or more of the activities mentioned above.

# 9. Discussions about suitable organisational forms

## 9.1 How to organise sponsored research work

The following discussion about how to organise research projects is based on the assumption that DFI adopts a programme for research and development based on the ideas put forth in chapter 8 and that cooperation between DFI sponsored projects and other related research work in information technology is to be encouraged.

We now consider two different forms for DFI to sponsor research projects:

A. The research budget is spent on a number of disparate projects.

Some of the characteristics of such an arrangement are:
- every single project is preceeded by an application, that is reviewed by DFI staff and possibly refereed by experts in the subject area, whereon the application is either granted or rejected;
- no deliberate cooperation between sponsored projects exists;
- a project is usually not sponsored for more than one year at a time;
- few project leaders dare set their projects rolling before DFI's financial support is secured.

B. The research budget is divided between one or more research teams.

In this case:
- each team is sponsored for at least three years at a time;
- no special applications are needed during this period. The activities of each team is decided in cooperation between DFI and the team;
- each team should have a team leader responsible for the overall economics of the team and, if possible, actively partaking in the research activities of the team.

Alternative A is close to the one now used by DFI while B is similar to the organisation used within the "Programme for Research in Information Processing and Computing Science" (see section 6.1).

If DFI decides to try to organise a long range programme for research in information technology according to alternative B, this would mean that a lot of work will have to be done initially to work out a programme, that covers the most important areas of IS&T. The outcome of this work is, of course, all the more important as a considerable part of the DFI research budget will then be tied to certain research teams for a long time, while under alternative A it would be possible to change the overall direction, if any, of the research programme

on short notice.

Another initial difficulty with alternative B is to find competent persons to be members of the research teams and above all to lead them and to decide on a suitable organisational setting for the research teams.

In spite of these difficulties, which, undoubtedly, can all be solved, we believe that DFI should try to form research teams as the advantages of such an organisation compared to the one now used are many:

1. once the research teams have been formed the work of DFI to administer the research activities should diminish, making possible increased efforts on other activities;

2. at least one team, specializing in IS&T, would exist in Sweden to be consulted by DFI in technical matters;

3. the whole area of research is one where expensive equipment such as computers of various sizes, secondary memory devices, and terminals is a necessity. Under alternative B the sharing of this equipment between many teams is made easier, hence cutting down the overall costs. This point is very important in the face of the expected development of new hardware devices;

4. the research teams will hopefully constitute such a fruitful environment that many competent people will be attracted to them in order to work with the problems of IS&T;

5. the work will be more efficient since the above-mentioned time lag under alternative A between idea and start of project does not apply;

6. very importantly, research work is by its nature a long range activity. It is seldom possible to achieve any results in one year only;

7. and most importantly of all, looking at the possibility of cooperation between DFI sponsored projects and other related projects in Sweden and/or abroad, one should remember that such cooperation always includes some form of negotiations in its initial stage, when the form for a joint project is decided. Such negotiations are usually not meaningful if the conditions are unknown, or not known for more than a limited period of time, by one of the parties. If alternative A is used and cooperation is to take place DFI would have to take part in three-way negotiations between the parties. We do not believe this to be an acceptable situation and find alternative B necessary if cooperation is desirable.

We do not investigate here how these teams should be organised, e.g as delegations, as foundations, as part of a university department or a research division in an industrial company. Some of the above problems are described in <SOU 1977;52>, where

for instance research institutes are compared to university departments.

For reasons of clarity we will, however, give some examples of possible organisational forms for the teams and of possible relations between the teams, and between the teams and other research groups. It should be noted that the following is to be considered only as examples and does not constitute any recommendation.

Team A:
This team is divided into three projects with the following activities:

- to do basic research on data base models for handling textual information;
- to study the problems of how to interface different system types both theoretically and through the implementation of prototypes;
- to follow the development of new hardware, e.g. videodiscs and data base machines.

The team should host a suitably configured computer system.

Team B:
This team does basic research in applications of artificial intelligence techiques to information retrieval systems.

Team C:
This team does some basic research in and development of distributed data base techniques for information retrieval systems. Members of the team participate in the relevant committees devoted to the standardization of the field and follow the development of commercially available software for distributed data bases.

Team D:
This team develops small systems according to given specifications (see section 3.4).

The teams could all cooperate with each other and with outside groups in various ways. Teams A and B could cooperate in that prototype data base systems developed within team A could be used by team B. Team C could cooperate with the SUNET project and also with Televerket and SCB. Team D cooperates with Office-85 and with various STU-financed research projects. All four teams could share much of the needed equipment.

## 9.2 How to get useful products from research efforts

Up till now we have rarely used the words research and development together. We believe that research projects and development projects should be kept apart in this context because different results are expected from them and they should therefore be organised in different manners.

Here we separate a research project from a development project

in that a research project is expected to
- increase the knowledge about the problems involved in a certain subject area of IS&T and/or
- result in a prototype version of a new product that might later (through a development project) well be used in an operational environment.

A development project on the other hand is initiated to develop a new product starting from a given specification of the characteristics of the product.

No clear line can of course be drawn between the two project types, but the following three examples will give an idea of how to proceed from a prototype developed within a research project to a useful product and also of how important it is not to mix the two project types initially. The word maintenance in this context means the correction of errors in a computer program so that it works according to the specifications given and the will and ability to effectuate some improvements of the program in response to requests from the users. It is a wellknown fact that for a system to be used in an operational environment it is necessary that it is properly maintained, in the sense of the word used here. The question, then, is: How should a prototype version of a computer system be developed into a useful product that is regularly maintained and who should pay for the maintenance?

### 9.2.1 The DATALAB example

DATALAB <Arnborg 1979> started as a research project named "Measurement data base methodology" and has been carried out at FOA since 1975, with the objective to evaluate and later to utilize the relational data model as a basis for analysis of scientific data, where the research worker should be able to use the program and manipulate the data without having to consult a system specialist. In 1979 a design specification was prepared by the research group and in 1980 a decision was taken to develop a product based on this specification, with the objective that DATALAB should be extensively used on the DEC10 machine at the Stockholm Computing Centre (QZ) and that it should also be possible to use DATALAB on other types of machines both in Sweden and abroad. Some important characteristics of this development work are:
- the development is sponsored by four parties in cooperation;
  FOA, QZ, the National Central Bureau of Statistics (SCB), and STU;
- the programming is made by the software house ENEA DATA on a contract from FOA;
- a test phase is planned to last for two years beginning in July 1981. SCB is to be responsible for and carry out the test of DATALAB;
- after the test phase QZ will take over the ownership of DATALAB and thus be responsible for the operation, distribution, maintenance, and further development of the product. The work involved in this is to be financed through

other people than the original sponsors of DATALAB paying a special fee for the use of the program.

### 9.2.2 The 3RIP example

The 3RIP project, sponsored by DFI's predecessor SINFDOK, started in 1973 as a "research and development" project. In 1977 a prototype of the entire system existed, which was not then ready for regular use. This prototype was tested and found useful after several revisions had been made. The research team formed the Paralog company and took over the rights to the product and developed it further, partly sponsored by SINFDOK. In 1979 the company began selling the system, with the agreement that a proportion of the gross income from the system is to be paid back to DFI. As a result of this decision by Paralog 3RIP is now commercially available on several installations in Europe and the US, Swedish users within DFI's area of competence may use the system with no charge to Paralog, and the maintenance of the system does not cost DFI anything.

This example shows how important it is to state clearly what the main objective of a given project is. In this case it was to produce an operational system, but if Paralog had not been created and taken over 3RIP, the problems of regular maintenance for the system would not have been solved - indeed, could not have been solved within the framework of DFI - and the whole project might have been regarded as a complete failure.

### 9.2.3 The LIBRIS example

The development of a library information system (LIBRIS) grew out of a study, commissioned by SAFAD in 1967, of the inner organisation of the Swedish governmental research libraries. The system development started in 1970/71 with the specification of an integrated libray information system. The scope and character indicated that LIBRIS would be a "research and development" project.

The programming of the system was done by a consulting firm with no prior experience in library automation. In 1972 a first version of the cataloging functions of LIBRIS was ready and could be used experimentally. LIBRIS was changed significantly in 1974 due to the introduction of new cataloging rules and the system was renamed LIBRIS II. For LIBRIS II the number of participating libraries was increased to ten and a desicion was taken to abandon previous plans for other functions than cataloging support within LIBRIS. However, the project was still considered to be "experimental", which meant that the system was maintained on an ad-hoc basis and that SAFAD only had responsibility for the operations of LIBRIS on a "project" basis.

In 1977 the LIBRIS activities were evaluated by the Riksdagens Revisorer <Gaje S. 1977>, who found that the system had not

been successful. Apart from the high cost of the project it was found that the quality of the service was low and that there was no clear administrative procedures for resolving occurring problems. "Low software quality" and "unclear responsibilities" were cited as possible causes. At that time the contract for the operation of LIBRIS was about to end and a permanent organisation was looked for. A decision was made to maintain a continuous LIBRIS operation during the change to a permanent situation and therefore the existing contract with SAFAD was prolonged. It was also decided to develop a new version of LIBRIS programmed in "standardized" COBOL to be run on an IBM computer. During this conversion no significant improvements were made to the system.

At present LIBRIS is installed at the governement owned computer centre DAFA.

The LIBRIS example, then, is a story of a project where research and development aspects have been constantly in conflict. The mechanisms necessary to control a development project have not ben used, probably because the project has been regarded as an experimental development. Furthermore, in the latest phases of the project the operational commitment has made it impossible to perform any experimental activities involving the system.

## 9.3 Concluding remarks on the differences between project types

There are, of course, many more models than the ones illustrated by our examples that could be used to disseminate research results to industry and to alleviate the creation of commercially available products. However, some general conclusions can be drawn from the examples, which should be considered when suitable organisational forms for research and development in IS&T are discussed:

- the sponsoring body should, when possible, separate between what is to be considered a research project and a development project;
- various organisational forms could be considered when a prototype is to be developed into an operational product; for instance, a recognized software house could be used as in the DATALAB example. It is not advisable, however, to involve a research team in such a development project;
- if one of the overall objectives of DFI is to develop products within the IS&T field then the cost of the development phase must not be underestimated; it should be noted from the DATALAB example that the amount of work needed to set up an organisation for efficient testing of a product and to carry out these tests is quite considerable, and also that competent groups do not come cheap;
- the maintenance problem must be given much consideration when a decision about development projects is made. Can a company be made to sell the product commercially - as in example B - or does an organisation exist with a competent group which could take over the product after it is developed

and be responsible for the maintenance at a reasonable cost - as we believe ought to have happened in the LIBRIS example?

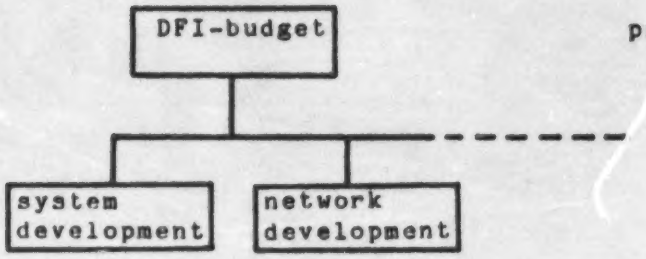
## 10. Suggestions

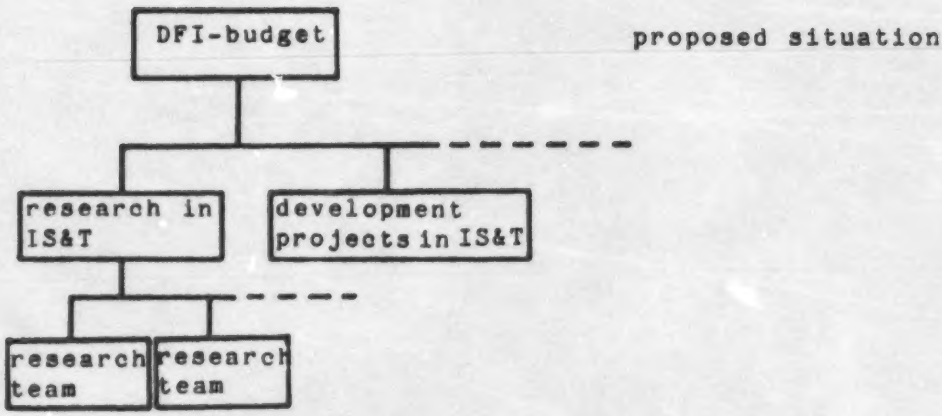
In this chapter our main proposals for a future DFI policy are summarized.

Suggestion 1
DFI should work out a programme for research and development in IS&T based on the ideas put forth in chapter 8.

The discussion in chapter 9 leads us to make the following four suggestions:

Suggestion 2
The parts of the DFI budget presently used for "system development" and "network development" should be re-channelled in accordance with this figure:



Suggestion 3
The total amount spent on research and development in IS&T should be split between the research activities and development projects in a way that reflects DFI's overall objectives.

Suggestion 4
The research work should be carried out by one or more research teams where
- each team is sponsored for at least three years at a time;
- the activities (projects) within each team is decided in cooperation between DFI and each team.

Suggestion 5
The only objective of a development project should be to develop complete products. If the developed product includes a computer program, this program should be maintained during the lifetime of the product. Development projects should be carried out after application to DFI in the same manner as now.

In accordance with the contents of chapters 6 and 7 we propose:

Suggestion 6
Cooperation between projects under the DFI programme and related projects elsewhere should be strongly encouraged.

## 11. A schedule of time and action

To realize a programme like the one proposed in chapter 10 the following action must now be taken by DFI:

1. Elaborate the ideas in chapter 8 to form a complete proposal for a programme for research and development in IS&T and rank the different activities according to their importance.

2. Define appropriate parts of such a programme that could possibly form the activities of one or more research teams.

3. Investigate how the proposed teams could best be organised.

4. Try to find interested persons to be members of such teams and in particular to serve as team leaders. Start discussing in detail what should be the activities of each team.

5. Investigate what ongoing and planned projects would be most valuable to cooperate with and start negotiations about such cooperation.

The above mentioned programme could be in effect by July 1982, assuming that DFI gets governemental support for it. For the budget year 1981/1982 we suggest that priority is given to applications coinciding with the proposed programme and above all to those carried out by persons or groups willing to participate in research team activities in the future.

## App. A: Research groups working with information processing in Sweden

Chalmers University of Technology and The University of Gothenburg, Department of Computer Science: One group has two main interests, programming methodology and computer system architecture.

The other group is mainly concerned with systems development methodology and data base design methodology.

The Royal Institute of Technology, Stockholm, Department of Computer Systems: This group is concerned with computer system architecture, microprogramming, programming methodology and operating system theory.

The Royal Institute of Technology, Stockholm, Department of Numerical Analysis and Computer Science: This group is concerned with tools for mathematical problem solving, graphical data processing, tools for text handling, and data base theory.

The Royal Institute of Technology and The University of Stockholm, Department of Information Processing and Computer Science: This department has various research groups, one group is concerned with systems development problems (ISAC), another group with computer aided learning (PRINCESS), and a third group with data base theory, simulation and programming methodology (CADIS).

The Linköping Institute of Technology, Department of Computer Science: This group is mainly concerned with programming systems and tools for programming and systems development. There are also some activities concerned with theoretical aspects of programming.

The Linköping Institute of Technology, Department of Systems Technique: Concerned with computer system architecture and computer aided design of integrated circuits. There is a quite comprehensive programme for picture processing.

The Institute of Technology of Lund, Department of Digital and Computer Technology: This group is concerned with computer architecture and digital technique. They are also investigating the area of programming new computer architecture.

Institute of Technology of Lund, Department of Automatic Control: The primary interest for this group in the area of information processing is languages for concurrent processes and principles behind such languages.

University of Uppsala, Department of Computer Technology: Concerned with methods and tools for design of micro processor systems, both hardware and software. They are also designing and producing computer network nodes.

University of Uppsala, Laboratory for Computer Science: Concerned with programming methodology, data bases and program manipulation.

University of Uppsala, Department of Administrative Data Processing: This group is concerned with data base design, systems development methodology, programming methodology and applications for small personal computers.

## App. B: Persons with whom contact has been taken

| Contacts: | Organisations: |
|---|---|
| Raj Reddy | Carnegie Mellon University<br>Department of Computer Science<br>Pittsburgh, PA 15203 |
| Paul Masters | Centigram Corporation<br>155A Moffett Park Drive, Suite 108<br>Sunnyvale, CA 94086 |
| Jan Carlsson | Data- och Elektronikkommitten<br>(Comp. and Electronics Commission))<br>111 51 Stockholm |
| Staffan Löf | Försvarets Forskningsanstalt<br>(Nat. Defence Research Inst.)<br>90 182 Umeå |
| Jacob Palme | Försvarets Forskningsanstalt<br>(Nat. Defence Research Inst.)<br>104 50 Stockholm |
| Ronald Kay | IBM Research Laboratories<br>5600 Cottal Rd<br>San Jose, CA 95193 |
| Hans Nilsson | Karolinska Institutet<br>Systemgruppen<br>Box 60400, 104 01 Stockholm |
| Hans Karlgren | KVAL<br>(Inst. for Inf. Sci.)<br>104 65 Stockholm |
| Charles Goldstein | Lister Hill Center for Biomedical<br>Communication - NLM<br>Bethesda, MD 20014 |
| Tamas Doszkocs | National Library of Medicin<br>Bethesda, MD 20014 |
| Ed Weiss | National Science Foundation<br>Washington, D.C. 20550 |
| Naomi Sager | New York University<br>Linguistic String Project<br>251 Mercer St<br>New York, NY 10012 |
| Frederick Hayes-Roth | Rand Corporation<br>1700 Main St<br>Santa Monica, CA 90406 |

| | |
|---|---|
| Rolf Gavare | Språkdata<br>(Gothenburg Univ.<br>Dep. of Comp. Ling.)<br>Norra Allegatan 6, 413 01 Göteborg |
| Patricia Whiting-O'Keefe<br>Gary Hendrix | SRI International<br>333 Ravenswood Ave<br>Menlo Park, CA 94025 |
| Gerald Caplan<br>Mike Genesereth | Stanford University<br>Department of Computer Science<br>Stanford, Ca 94305 |
| Krister Gustavson<br>Börje Holmborn | Statskontoret<br>(Swedish Agency for Admin. Dev.)<br>Box 34107, 100 26 Stockholm |
| Lars Kahn | Styrelsen för Teknisk Utveckling<br>(Nat. Board for Tech. Dev.)<br>100 72 Stockholm |
| Alex Eckerström | Televerket<br>(Nat. Telecomm. Admin.)<br>123 86 Farsta |
| Werner Schneider | UDAC<br>(Upsala Univ. Comp. Centre)<br>Box 2103, 750 02 Uppsala |
| Robert Wilensky | University of California, Berkeley<br>Department of Computer Science<br>Berkeley, CA 94720 |
| Nolan Pope | University of Florida Library<br>Gainesville, FL |

List of references:

<Ada 1980>:
Ada Reference Manual. Proposed Standard Document. United States Department of Defense. July 1980. (Published by FOA, Stocholm, Sweden).

<AIC>:
The INTELLECT Query System is marketed by Artificial Intelligence Corporation, 200 Fifth Ave, Waltham, MA 02254.

<Altemark S. 1980>:
A Portable and User Oriented Data Base Management System and its Application in the Medical Field. MEDINFO 80, Linberg/Kaihara Eds. North Holland Publishing Company, 1980.

<Ammon G.J. 1979>:
Wideband optical disc data recorder systems. In SPIE Vol.200 Laser Recording and Information Handling, 1979. pp 64-72.

<Arnborg S., Elvers E., Svensson P. 1979>:
Design Specification for DATALAB-A System for Data Analysis on the Relational Model of Data. FOA Report C20362-D8.

<Ashford J. and Martin D. 1979>:
The Future for Free Text Systems. Computer Bulletin, September 1979. p 20-22.

<Bartolini R.A. et al. 1978>:
Optical Disc Systems Emerge. IEEE Spectrum, August 1978. pp 20-28.

<Berra B. 1977>:
Data Base Machines. In SIGIR Forum. Vol. XII, No. 3. pp 5-23.

<Berra B. 1980>:
Interview made with Mr. Berra in Computerworld 1980-12-15 by John Whitmarsh.

<Bihan J. et al. 1980>:
SIRIUS : A French Nationwide Project on Distributed Data Bases. In French-Swedish Symposium. Subject B: Telematics. Stockholm, November 1980.

<Bourne C. 1980>:
On-line Systems: History, Technology, and Economics. JASIS, May 1980, pp 155-160.

<Centigram>:
The MIKE Voice Recognition and Response System is marketed by Centigram Corp., 155A Moffett Park Drive, Suite 108, Sunnyvale, CA 94086.

<Cherry S. S. 1981>:
The Moving Finger "Accesses." American Libraries, January 1981 (reprint).

<CLASSIC>:
Marketed by Cincinnati Electronics, Cincinnati, OH

<CLSI>:
Marketed by CL Systems, Inc., 81 Norwood Ave., Newtonville, MA 02160.

<Copeland G. 1980>:
What if mass storage were free?. In The Papers of the Fifth Workshop on Computer Architecture for Non-numeric Processing. March 1980. pp 1-7. (Also referred to as SIGIR Vol.XV No. 2).

<Codd E.F. 1970>:
A Relational Model of Data for Large Shared Data Banks.
CACM, Volume 13, No. 6. pp 377-387.

<Codd E.F. 1981>:
SQL/DS What it means. In Computerworld 1981-02-16. p 27.

<Dagens Industri 1981>:
Interview made with Håkan Ledin from LM Ericsson by B G Wennersten. Dagens Industri, 1981-01-22.

<Dataphase>:
Dataphase Systems, 3770 Broadway, Kansas City, MO 64111.

<Davies D.W. et al. 1979>:
Computer Networks and their Protocols.Wiley, 1979.

<Doszkocs T.E. and Rapp B.A. 1979>:
Searching MEDLINE in English: a prototype user interface with natural language query, ranked output, and relevance feedback. Proc. ASIS Annual Meeting, 1979, 16, pp. 131-139.

<Electronic Business 1980>:
Markets of the '80s, Office of the future to use integrated text/data processing. Electronic Business, February 1980.

<Ellis C. and Nutt G. 1980>:
Office Information Systems and Computer Science. Computing Surveys, March 1980.

<Foster C.C. 1976>:
Content Addressable Parallel Processors. Van Nostrand, 1976.

<Frost&Sullivan 1981>:
Future Office Systems, Equipment and Associated Software Markets. Report from the Frost&Sullivan Company in New York 1981.

<Frost&Sullivan 1980>:
The European Data Base Market. Report No. E264 from the Frost&Sullivan Company in New York, 1980.

<Gaje S. 1977>:
Det datorbaserade informationssystemet för vetenskapliga bibliotek-LIBRIS. Granskningspromemoria nr 10/1977 från Riksdagens Revisorer. (In Swedish)

<Geac>:
Geac Canada Ltd., 350 Steechase Road W., Markham, Ont. L3R 1B3.

<Golstein C.M. 1979>:
Videodisc Mass Storage Program. Memorandum to the Lister Hill Center Scientific Review Committee. October 2, 1979.

<Goldstein C.M. and Dick R.S. 1980>:
The LHC Integrated Library System, Version 1.0. NLM News, vol XXXV, nr 9, pp. 1-3.

<Goldstein C.M. and Ford W.H. 1979>:
Program in Computer-based Educational Systems. A report available from the Lister Hill National Center for Biomedical Communications.

<Hendrix G.G. 1977>:
The LIFER Manual: A Guide to Building Practical Natural Language Interfaces. Technical Note 138, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, CA, February 1977.

<Hendrix G.G. and Haas N. 1980>:
An Approach to Acquiring and Applying Knowledge. Technical Note 227, Artificial Intelligence Center, Computer Science and Technology Division, SRI International, 333 Ravenswood Ave, Menlo Park, CA 94025, November 1980.

<Hiltz S.R. and Turoff M.>:
The Network Nation: Human Communication Via Computer. Addison-Wesley, 1978.

<HSTS 1980>:
HSTS High Speed Text Search System. Sales brochure from Operating Systems inc. 1980.

<Hägglund S. 1980>:
Contributions to the Development of Methods and Tools for Interactive Design of Applications Software. Linköping Studies in Science and Technology Dissertations No 54. Linköping University, 1980.

<Höglund L. and Persson O. 1981>:
(A Study for DFI, to be published in february 1981.)

<ISOK 1980>:
Ramprogram ISOK (Informationsbehandling för Samhällsplanering och Kommunalteknik): Plan för systemutveckling och systeminförande. Kommunförbundet 1980. (In Swedish.)

<Janning M. et al. 1979>:
Introduction to Associative Databases and the CS4 System. Studentlitteratur, Lund, 1979. (In Swedish).

<Karlgren H. and Walker D.>:
The POLYTEXT System - A New Design for a Text Retrieval System. Technical Report, Artificial Intelligence Center, SRI International, Menlo Park, CA, 1980.

<Kurzweil>:
The Kurzweil Data Entry Machine (KDEM) is marketed by Kurzweil Computer Products, 33 Cambridge Parkway, Cambridge, MA 02142.

<Lancaster 1980>:
On-line systems in the Communication Process: Projections.
JASIS May 1980. pp 193-200.

<Lancaster F.W. et al. 1973>:
Information Retrieval On-line. Wiley, 1973.

<Larsson R. 1980>:
On Implementing the Common Command Language within 3RIP. Royal Institute of Technology Library, Stockholm, April 1980. (Report TRITA-LIB-4067.)

<Lea W.A., Ed. 1980>:
Trends in Speech Recognition. Prentice-Hall, Englewood Cliffs, N.J. 1980.

<Lindquist M.G. 1974>:
Automated Library Networks in the USA and Canada. Royal Institute of Technology Library, Stockholm, August 1974. (Report TRITA-LIB-1058).

<Lindquist M.G. 1980>:
3RIP-COM: Integrating Information Retrieval and Computerized Conferencing. Proc. 43rd ASIS Annual Meeting, Vol 17, 1980.

<Lindsay R.K. et al 1980>:
Applications of Artificial Intelligence to Chemistry: The DENDRAL Project. McGraw-Hill, New York, 1980.

<Löf S. 1980>:
The POLYTEXT/ARBIT Demonstration System. FOA 4 report C 40127-M7. September 1980.

<Meadow C.T. 1979>:
The Computer as a Search Intermediary. Online, 1979, 3 (3), pp. 54-59.

<Medress M.F. et al. 1978>:
Speech Understanding Systems: Report of a Steering Commitee. Artificial Intelligence, Vol. 9, 1978. pp 307-316.

<MIMIS 1980>
MIMIS. Manuskriptversion 3 801215. UDAC, Uppsala. (A system description in manuscript. In Swedish.)

<Nora S. and Minc A. 1980>:
The Computerisation of Society. A Report to the President of France. MIT Press, 1980. ISBN 0-262-14031-4.

<NSF 1979>:
Information Science and Technology. National Science Foundation Program Report, vol 3, nr 5.

<OCLC>:
OCLC, Inc., 1125 Kinnear Rd., Columbus, OH 43212.

<Online Computer Systems, Inc.>:
20010 Century Blvd., Germantown, MD 20767.

<PACTEL 1979>
Distributed Database Technology. A report prepared by (PA Computera and Telecommunications). NCC Publications, 1979. ISBN 0-85012-226-0

<Palme J. 1980-11-22>:
GILT-projektet: Interconnection Between Local Text Systems. Message number 182035 from the COM system at QZ.

<Palme J. November 1980>:
The COM Computerised Teleconferencing System. In French-Swedish Symposium. Subject B: Telematics. Stockholm, November 1980.

<Palme J. December 1980>:
Erfarenheter av användning av telekonferenssystemet KOM. FOA report C 10166-M6(H9.) December 1980. (In Swedish.)

<Paralog 1980>:
3RIP An Introduction. Paralog, Stockholm, July 1980.

<Regazzi J. et al. 1980>:
On-line systems of Disciplines and Speciality Areas in Science and Technology. JASIS, May 1980. pp. 161-170.

<Resnikoff H. 1980>:
Information Science and Technology at the National Science Foundation. Information Processing & Management. Vol 16. pp 243-250.

<RDF 1980>:
Riksdataförbundet. Versamhetsberättelse 1979. (In Swedish.)

<Riksdataförbundet 1980>:
80-talet på en bricka. Published by The Swedish Federation of Data Processing Users 1980. (In Swedish.)

<Robinson B.M. 1980>:
Cooperation and Competition among Library Networks.JASIS, November 1980, pp. 413-424.

<SAFAD 1980>:
Utvecklingsplan 1980. Statskontoret 1980-08-11. (In Swedish.)

<SAFAD 1980:13>:
Handlingsprogram för metodstöd på ADB-området. Statskontoret, Report 1980:13. (In Swedish.)

<Sager N. 1978>:
Natural Language Information Formatting: The Automatic Conversion of Texts to a Structured Data Base. Advances in Computers, Vol 17, Academic Press, New York, 1978.

<Samuelson K. et al 1980>:
Informatic Communication and Multi-way Video. Tape recording of special session 4, ASIS Annual Meeting 1980. Minute Tape, Van Nuys, CA 91405, 1980.

<Sandberg A. 1980>:
SCENARIO över framtida FOA-användning av KOM - 1985. FOA report C 10166-M6(H9), pp. 107-130. December 1980. (In Swedish.)

<Sandewall E. 1978>:
Programming in the Interactive Environment: The LISP Experience. Computing Surveys, March 1978.

<Sandewall E. 1979>:
Why Superroutines are Different from Subroutines. Research Report LiTH-MAT-R-79-28 from the Informatics Laboratory, Linköping University.

<Shortliffe E.H. 1976>:
Computer-Based Medical consultations: Mycin. American Elsevier, New York, 1976.

<Siemens 1978>:
The CONDOR project is run as a joint venture by Siemens and the Bundesministerium fur Forschung und Technologie. A number of research reports and articles printed together under the name "CONDOR: Modell eines integrierten Datenbank- und Informationssystem" are available directly from Siemens AG, Munich.

<Smalley D.A. et al. 1980>:
Linking the Bibliographic Utilities: Benefits and Costs. Battelle Columbus Laboratories, Columbus, OH, September 1980.

<SOF 1980>:
Sök Och Finn (SOF) på TDS systemet. Program-makarna, Stockholm 1980. (In Swedish)

<SOU 1977:52>:
Forskningspolitik. Betänkande av forskningsrådsutredningen. SOU 1977:52. (In Swedish)

<SOU 1980:26>:
Mot bättre vetande. Förslag från uppslagsverkskommitén. SOU 1980:26. (In Swedish).

<STU 1980>:
Programme for Research in Information Processing and Computer Science. Kahn L. Ed. National Board for Technical Development, Stockholm 1980. (STU-Information nr 186-1980.)

<SYSTEX>:
SPIRIT SYSTEM: An Entirely Automatic Storage and Retrieval System. In French-Swedish Symposium. Subject B: Telematics. Stockholm, November 1980.

SYSTEME S.P.I.R.I.T.: Systeme Syntaxique et Probabiliste d'Indexation et de Recherche

d'Informations Textuelles. This is a report describing the system, which is also marketed by SYSTEX, 88 Allee du Lavoir, 91190 Gif-sur-Yvette.

<Tomberg A. 1980>:
The Development of Commercially Available Databases. Online Review, December 1979. pp 343-353.

<Wanger J. and Landau R. 1980>:
Nonbibliographic On-line Data Base Services. JASIS, May 1980. pp 171-180.

<Warren S. K. and Abbe D. 1979>:
A Conversational, Extensible Microcomputer Language. In SIGSMALL Newsletter, Vol. 5, No. 2, April 1979.

<Wiederhold G., Kaplan J.S., and Sagalowicz D. 1980>:
Research in Knowledge Base Management Systems. Research report available at the Dept. of Comp. Sci., Stanford University, Stanford, CA, 1980.

<Wilensky R. 1980>:
META-PLANNING: Representing and Using Knowledge about Planning in Problem Solving and Natural Language Understanding. Memorandum No. UCB/ERL M80/33, 5 August 1980, Electronics Research Laboratory, College of Engineering, University of California, Berkeley, CA.

<Wilensky R. and Yigal A. 1980>:
PHRAN: A Knowledge-Based Approach to Natural Language Analysis. Memorandum No. UCB/ERL M80/34, 12 August 1980, Electronics Research Laboratory, College of Engineering, University of California, Berkeley, CA.

<Williams P.W. 1980>:
The Use of Microelectronics to Assist Online Information Retrieval. Online Review, December 1980. pp 393-399.